


The Coin in Hand–Extended Version: Development and Validation of a Multicultural Performance Validity Test

Assessment
1–13
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1073191119864652
journals.sagepub.com/home/asm


Julia C. Daugherty^{1,2} , Luis Querido³, Nathalia Quiroz⁴, Diana Wang⁵,
Natalia Hidalgo-Ruzzante^{1,6}, Sandra Fernandes³, Miguel Pérez-García^{1,2},
Carlos Jose De los Reyes-Aragon⁴ , Rute Pires^{3,7}, and Eve Valera⁵

Abstract

The number of computerized and reliable performance validity tests are scarce. This study aims to address this issue by validating a free and computerized performance validity test: the Coin in Hand–Extended Version (CIH-EV). The CIH-EV test was administered in four countries (Colombia, Spain, Portugal, and the United States) and performance was compared with other commonly used validated tests. Results showed that the CIH-EV has at least 95% specificity and 62% sensitivity, and performance was highly correlated with scores on the Test of Memory Malingering, Victoria Symptom Validity Test, and Digit Span of the Wechsler Adult Intelligence Scale. There were no significant differences in scores across countries, suggesting that the CIH-EV performs similarly in a variety of cultures. Our findings suggest that the CIH-EV has the potential to serve as a valid validity test either alone or as a supplement to other commonly used validity tests.

Keywords

performance validity, simulation, feigning, cross-cultural, effort

Standard protocols in neuropsychological assessment call for the use of validity tests in order to control for variables such as lack of effort, or attempts to exaggerate or malingering symptoms (Heilbronner et al., 2009; Martin, Schroeder, Heinrichs, & Baade, 2015; Mittenberg, Patton, Canyock, & Condit, 2002). Some neuropsychological validity tests use a forced-choice response design in which the evaluatee is presented with a stimuli (such as an image or a series of digits), and later asked to recognize the same stimuli by choosing between two separate stimuli, one correct and one incorrect (Frederick & Speed, 2007). Forced-choice testing allows professionals to calculate below-chance response patterns (Frederick & Speed, 2007), scores below what would be expected when guessing or responding at random (Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012). Despite the ease with which forced-choice tests are performed by individuals with neurocognitive impairment, many people who wish to exaggerate symptoms perform with exceedingly low scores that individuals with cognitive impairment rarely reach (Schroeder, Twumasi-Ankrah, et al., 2012). The use of such tests allows professionals to measure whether diminished neuropsychological scores reflect the aforementioned variables or true cognitive impairment. Unfortunately, there are very few validity tests that are free, making these test less likely to be used in cases where there are limited financial resources. As such, the

present study aimed to validate a free performance validity test while evaluating the cultural equivalency of the measure across four countries: Spain, the United States, Portugal, and Colombia.

Although there are several performance validity tests in current use (Griffin, Glassmire, Henderson, & McCann, 1997; Kapur, 1994; Rey, 1964), more tests are needed for a number of reasons. First, the rather limited number of validity tests increases the probability of successful coaching and familiarization with instruments (Suhr & Gunstad, 2007).

¹The Mind, Brain and Behaviour Research Centre (CIMCYC), University of Granada, Granada, Spain

²Department of Personality, Evaluation, and Psychological Treatment, Universidad de Granada, Spain

³Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, Lisboa, Portugal

⁴Psychology Department, Universidad del Norte, Baranquilla, Colombia

⁵Department of Psychiatry, Harvard Medical School, Boston, United States

⁶Department of Evolutionary and School Psychology, Universidad de Granada, Spain

⁷CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, Lisboa, Portugal

Corresponding Author:

Natalia Hidalgo-Ruzzante, The Mind, Brain, and Behaviour Research Centre (CIMCYC), University of Granada, Campus Cartuja S/N, Granada 18071, Spain.
Email: nhidalgo@ugr.es

Furthermore, with the proliferation of information about validity tests on the web, coaching oneself on how to pass validity tests is made simpler. To avoid detection, some patients may educate themselves on validity testing through Internet searches or published journal articles and books (Suhr & Gunstad, 2007). Furthermore, attorneys may provide coaching to their clients prior to neuropsychological testing (Youngjohn, 1995). Increasing the number of validity tests permits a wider selection for clinicians to choose from so that they can (a) increase sensitivity by employing multiple validity measures in the battery (Chafetz, 2011; Larrabee, 2003; Meyers & Volbrecht, 2003; Sollman, Ranseen, & Berry, 2010) and (b) decrease the likelihood of familiarity with specific validity tests (Lippa, 2018).

Second, cross-study comparisons are made difficult by the lack of uniformity in instrumentation (Gershon et al., 2010; Nijdam-Jones & Rosenfeld, 2017). To these ends, there has been a growing initiative to provide computerized assessment tools. For example, the National Institutes of Health's (NIH) Blueprint for Neuroscience Research has developed the NIH Toolbox, a series of standardized neuropsychological assessments, to advance research using a "common currency" of validated measures (Gershon et al., 2010). All NIH toolbox instruments are free to approved researchers, eliminating the expensive costs that may inhibit the use of a wider range of assessments. Furthermore, the digitization of tests facilitates data collection on a large scale and the standardization of test administration and scoring. Computer-based testing permits the precise and immediate evaluation of performance on variables that may be more difficult to measure, such as time, without this technology. Despite these advancements, the NIH Toolbox does not include any validity tests. As such, there remains a great need to continue developing and validating novel methods of evaluating performance validity, especially in cross-cultural contexts.

We identified two validity tests available in the public domain; namely, the Coin in Hand Test (CIH; Kapur, 1994) and the Rey 15-Item Memory Test (Rey I: Rey, 1964; and Rey II: Griffin et al., 1997). Unfortunately for the Rey, several limitations and concerns have been found in the original test (e.g., nonfeigning participants were incorrectly identified as feigning; Reznick, 2005). Unfortunately for the Rey, several limitations and concerns have been found in the original test (e.g., nonfeigning participants were incorrectly identified as feigning, Reznick, 2005). To address some of these shortcomings, a redesigned version has been created to include a recognition trial and a combined recall and recognition score. While this modified version has shown some improvement (raising sensitivity from 47% to 50%; Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002), other studies report an alarming number of false positives in psychiatric outpatients (65%; Whitney, Hook, Steiner, Shepard, & Callaway, 2008) and forensic psychiatric populations (74.7%; Stimmel, Green, Belfi, & Klaver, 2012). In contrast,

the CIH does not share these same shortcomings. The original version of the CIH is a nondigitized test that was developed to determine if bedside patients demonstrate exaggerated memory complaints using a forced-choice selection (Kapur, 1994; Lezak et al., 2004). Kapur (1994) initially demonstrated that patients suffering from amnesia perform well on this task. Furthermore, he found that two patients who were at a greater probability of feigning performed close to chance level of simulation.

Subsequent research conducted uniquely on clinical populations (i.e., with no feigning or healthy control group) have found high specificity rates when using a cutoff of >1 among patients with dementia (88.1% to 89%; Rudman, Oyeboode, Jones, & Bentham, 2011; Schroeder, Peck, Buddin, Heinrichs, & Baade, 2012), brain injury (specificity, 95%; Hampson, Kemp, Coughlan, Moulin, & Bhakta, 2014), and epilepsy (93.8%; Hampson et al., 2014). Another study using clinical patients where specificity was not reported has likewise found that individuals with genuine intellectual deficiencies respond correctly to all trials (Colwell & Sjerven, 2005).

Other studies have selected an analog paradigm to compare healthy controls, experimental simulators, and clinical patients (i.e., amnesic, brain injury, or autoimmune disorders; Cochrane, Baker, & Meudell, 1998; Ferreira, Gomes, Moreira, Silva, & Cavaco, 2017; Hanley, Baker, & Ledson, 1999; Kelly, Baker, Broek, Jackson, & Humphries, 2005; Yeh et al., 2018). In these studies, the CIH was found to have a modest to high sensitivity (67% to 100%) among simulators and a high specificity (87.5% to 100%) among clinical populations. Of the aforementioned studies that have compared CIH performance with other commonly used performance validity testing (PVT), findings demonstrate the CIH to result in fewer false positives than the Test of Memory Malingering (TOMM; Rudman et al., 2011; Yeh et al. 2018) and the Rey 15-item (Rudman et al., 2011) among dementia patients. Finally, various studies have also demonstrated that CIH scores are independent of neurocognitive functioning, age, and education level (Ferreira et al., 2017; Schroeder, Peck, et al., 2012).

While these studies are useful in understanding the performance of this measure in both clinical populations and healthy controls, many studies do not report both the sensitivity and specificity, and none make cross-cultural comparisons to understand whether the test has cross-cultural biases. Furthermore, the original version of the CIH is not digitized and does not include various levels of perceived difficulty. Including these novel characteristics may help improve the original CIH by making it accessible to those who cannot afford other performance validity measures, and by increasing the precision and standardization through computerized data collection for variables that are more difficult to measure using traditional pen-and-paper methods. Furthermore, the sensitivity of validity tests has been

improved by increasing the apparent or actual difficulty of the task (Binder, 1990; Hiscock & Hiscock, 1989; Iverson, Franzen, & McCracken, 1994). In these cases, exaggeration can be detected by examining error rates across the “difficulty levels,” with more errors expected as the perceived or actual difficulty level increases (Chiu & Lee, 2002).

As such, the CIH test was selected to be computerized and modified to include multiple levels of perceived difficulty. Furthermore, instructions were developed in English, Spanish, and Portuguese with the objective of evaluating its cross-cultural application. The unique contribution of the present study to the current literature will be the validation of a digitized and multilingual instrument, transparency of sensitivity and specificity estimates for various cultures, and the use of a nonclinical and neurologically healthy population. While the clinical relevance of assessing university students is not readily evident, using this population is helpful in the validation of new PVT measures as it allows for the assessment of a base rate of PVT failure due to the fact that they have no apparent reason to perform poorly nor strong external incentives to do their best (An, Zakzanis, & Joordens, 2012). Gaining an approximation of little to no external incentives has been considered an essential criterion to creating a clean control group for validating PVTs (Slick, Sherman, & Iverson, 1999). To this end, two studies were undertaken to develop and study the psychometric properties of the Coin in Hand–Extended Version (CIH-EV), a modification of the CIH that addresses these shortcomings. In Study 1, we present the development of the new instrument as well as its psychometric properties (i.e., specificity, sensitivity, and convergent validity) in Spain. In Study 2, we present the cultural equivalency of the measure tested with participants from Spain, the United States, Portugal, and Colombia.

Study 1

The objective of Study 1 was to establish the psychometric properties of the CIH-EV in Spain.

Method

Participants. In total, 116 Spanish participants (20 male and 96 female) with an age range between 18 and 39 years ($M = 21.26$, $SD = 2.56$) were recruited from the undergraduate programs for psychology and nursing at the University of Granada (Spain).

Following the commonly used analog simulation design of previous studies assessing validity tests (Merten, Green, Henry Blaskewitz, & Brockhaus, 2005; Tydecks, Merten, & Gubbay, 2006), participants were assigned to one of two conditions (with or without instructions). As such, participants in the first condition ($n = 76$) were not informed about the three levels of difficulty, and were randomly

assigned to one of two groups: (a) the feigning group or (b) the control group. Feigning participants ($n = 36$) were asked to perform the CIH-EV as if they had suffered a brain injury (age range of 18-39 years; $M = 21.66$, $SD = 3.43$), while the control participants ($n = 40$) were advised to perform to the best of their abilities (age range of 18-29 years; $M = 21.1$, $SD = 2.46$). Participants in the second condition ($n = 40$) received instructions about the three increasing levels of difficulty on the CIH-EV and were also randomly assigned to either the feigning or control group. Twenty individuals were randomly assigned to the feigning group with an age range between 18 and 25 years ($M = 21.2$, $SD = 1.76$), and 20 to the control group with an age range between 18 and 24 years ($M = 20.85$, $SD = 1.42$). Inclusion criteria required participants to have Spanish as their native and dominant language. Participants were excluded from the study if they (a) had consumed psychotropic drugs or illicit substances within 24 hours prior to test administration and/or habitually, (b) had lost consciousness due to acquired hits to the head, or (c) were diagnosed of illnesses that could affect neuropsychological functioning.

Instruments

The Coin in Hand–Extended Version (Daugherty, Hidalgo-Ruzzante, & Pérez-García, 2017). The CIH-EV, adapted from the original CIH (Kapur, 1994), was developed as a digitized multiplatform tool that can be administered on tablets and personal computers. First, evaluatees pass through a series of screens on which they read the test instructions. On the first screen, instructions read,

When we are memorizing information and get distracted, we often forget this information. This test intends to measure how your memory can resist distraction. To this end, two hands will appear on the screen with a coin in one of them, such as these.

An image of two hands with a coin in one hand is presented. Once the evaluatee has read the instructions, he or she may continue onto the second screen which reads, “Next, both hands will close and the screen will turn black. Then we will ask you to count down.” A photo of two closed fists is shown on the screen, and the evaluatee may click “continue” to proceed with instructions. On the third screen, the instructions read, “The hands will reappear. When this happens, you must select the hand that previously held the coin.” On this screen, there is an image of two closed fists with a cursor selecting one of the two fists. Once the participant has selected “continue,” the task will begin. Three 10-trial levels of difficulty that vary in perceived difficulty are included in the CIH-EV, as the literature suggests that perceived difficulty improves the detection of feigning (Chiu & Lee, 2002). In the first level of difficulty, the participant must indicate which hand held the coin after counting down from the number 99 at a rate of one digit per

second for 10 seconds. The participant is guided by a visual countdown on the screen, and responses cannot be made until the full time has passed. Following the 10 trials of the first block, the following instructions for the second level of difficulty are presented on the screen, “Now you will begin another part of the test that is more DIFFICULT. Instead of counting down for 10 seconds, now you must count down for 15 seconds from the number 99 before choosing which hand had the coin.” Following the 10 trials of the second block, instructions for the third level of difficulty are presented as follows, “Now you will begin the most DIFFICULT part of the test. Instead of counting down for 15 seconds, now you must count down for 20 seconds from the number 999 before choosing which hand had the coin.” In the first version of the test, participants were not informed by evaluators about the increasing levels of difficulty prior to test administration and were only made aware by the on-screen instructions that appear immediately before the second and third block. In the second version, however, evaluators informed participants verbally about the three levels of difficulty prior to test administration. These instructions were given in addition to the instructions that were shown on-screen before the second and third block to all participants. Instructions were given in the second version to test whether sensitivity would improve by increasing the apparent difficulty of the task, as suggested by the literature (Binder, 1990; Hiscock & Hiscock, 1989; Iverson et al., 1994). Dependent variables include response time and correct hits for each level of difficulty as well as the combined total. The CIH-EV can be administered on tablets and computers as well as on all types of operating systems (i.e., Windows, Mac, etc.). The test can either be run online via the instrument’s web address (<https://projectbelieve.info/en/professionals/>) or after downloading the application from the app store. While the current version requires Internet connection, future versions will be available for download and offline use. Access must first be granted through the aforementioned webpage, which requires that users indicate their intended use of the instrument as well as professional credentials as a neuropsychologist or psychologist. This access is strictly monitored by the authors at the University of Granada (Spain).

The Test of Memory Malinger (Tombaugh, 1996). The TOMM is a performance validity test that is composed of 50 visual memory items, and two trial phases for learning and evaluation. In the learning trial, the participant views 50 line drawings for 3 seconds each. Next, in the evaluation phase, the participant is shown one of the previous drawings next to a new image, and asked to indicate which of the two images he or she had seen previously. In the second trial, the same process is repeated with the same drawings but in a different order of presentation. Two different cutoff scores may be applied to differentiate between scores that are affected by traumatic brain injury or neuropsychological

impairment, and those that are a result of feigning. The TOMM demonstrates high test–retest reliability for the two trials ($\alpha = .94$ and $.95$, for trials 1 and 2, respectively). With a cutoff score of 45 the TOMM demonstrates a sensitivity of 86% and a specificity of 100% in Spanish-speaking samples (Vilar-López et al., 2007).

The Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997). The VSVT is a computerized forced-choice validity test used to assess the exaggeration or simulation of cognitive impairment. It includes a total of 48 items that are presented in three blocks of 16 items each. In the first block, the participant is shown a number sequence, which then disappears for 5 seconds. Following the 5 second delay, he or she is presented with two different number sequences and must choose which one is identical to the one that first appeared. In the following blocks, the same procedure is followed but using 10 and 15 second delays. The total number of correct responses and latency for each block is scored. With a cutoff score of 44, the VSVT demonstrates a sensitivity of 97.1% and a specificity of 100% in Spaniards (Vilar-López et al., 2007).

Procedure. The study was carried out in the Mind, Brain, and Behavior Research Center at the University of Granada (CIMCYC-UGR). All participants signed an informed consent document before participating. In addition to the researcher’s contact information, the informed consent included information about confidentiality for the obtained data in accordance with the Organic Law 15/1999 for the Protection of Personal Data in Spain. The project was approved by the University of Granada’s ethics committee before testing.

All participants were administered the CIH-EV in addition to the two other performance validity tests (TOMM and VSVT) in a randomized order. The CIH-EV was administered to all participants using the same iPad model (iPad Pro 9.7 inch) with 2048-by-1536-pixel resolution at 264 ppi. Due to the finding that test sensitivity may be improved by increasing the apparent or actual difficulty of the task (Binder, 1990; Hiscock & Hiscock, 1989; Iverson et al., 1994), the influence of having received instructions about the difficulty levels on the CIH-EV was assessed. Students in the first condition were not given specific instructions about the difficulty levels, while participants in the second condition received instructions that the test would increase in difficulty over the three blocks. In both studies, half of the participants were told to perform to the best of their abilities (control group), while the other half (feigning group) was given a vignette in which they were told to simulate cognitive impairment following a car accident in order to receive economic compensation (Suhr & Gunstad, 2000).

Design and Data Analyses. The first objective of the study was to determine whether there were significant

Table 1. Reaction Time and Hits in the Feigning and Control Groups With and Without Instructions.

	No instructions				Instructions			
	Block 1	Block 2	Block 3	F	Block 1	Block 2	Block 3	F
Correct responses, <i>M</i> (<i>SD/SE</i>)								
Control	9.97 (0.16/0.26)	9.94 (0.22/0.25)	9.94 (0.22/0.23)	a, b, c	9.8 (0.69/0.35)	9.8 (0.69/0.38)	9.75 (0.91/0.33)	a, b, c
Feigning	6.54 (2.46/0.29)	5.9 (2.29/0.27)	5.4 (2.17/0.25)		5.89 (2.13/0.36)	5.1 (2.33/0.39)	4.89 (1.94/0.34)	
Delay, <i>M</i> (<i>SD/SE</i>)								
Control	1556 (430/163.79)	1603 (450/181.61)	1577 (420/196.32)	—, —, —	1289.72 (224/388.00)	1354 (229/437.73)	1250 (184/422.97)	—, —, —
Feigning	2765 (1427/175.42)	3003 (1603/194.50)	2773 (1740/210.26)		2331 (2443/388.00)	2722 (2758/437.73)	2862 (2668/422.97)	

Note. Correct responses = number of correct responses; delay = delay in response in milliseconds; SE = standard error; a = significant main effect for difficulty at $p < .005$; b = significant main effect for group at $p < .005$; c = significant main effect for interaction at $p < .005$.

differences between the three levels of difficulty. To do this, a two factorial 3×2 analysis of variance was conducted with difficulty level (Levels 1 through 3) and group (feigning vs. control) as independent variables, and hits and response time as dependent variables. The same analyses were conducted for the groups that did and did not receive information about the difficulty levels. Finally, an independent samples *t* test was completed with the same groups for total hits (the sum of the hits) and mean response time across difficulty levels. For the second objective, appropriate cutoff scores for hits and time were established using Receiver Operating Characteristic analyses for each trial as well as total trials as recommended (90% specificity and 50% sensitivity; Boone, Victor, Wen, Razani, & Ponton, 2007; Dean, Victor, Boone, & Arnold, 2008; Larrabee, 2012; Larrabee, Greiffenstein, Greve, & Bianchini, 2007; Morgan & Sweet, 2009; Sugarman & Axelrod, 2015). Using the new cutoff, a chi-square analysis was conducted to measure the percentage of false positives and negatives. Furthermore, a Spearman correlation was conducted to test the CIH-EV's convergent validity by comparing it with other commonly used performance validity tests in Spain (TOMM and VSVT). Last, a kappa chi-square analysis was completed to test for the degree of congruency between the CIH-EV, TOMM, and VSVT in detecting simulation.

Results

Assessment of Instructions and Levels of Difficulty. There were no differences in CIH-EV performance (as measured by total hits and total response time) by the order in which the CIH-EV was administered. This was true for the control group, the feigning group, and both groups combined.

For the group that did not receive information about levels of difficulty, there was a main effect for difficulty

level (Block 1 to 3), $F(2, 69) = 13.54, p < .001, \eta_p^2 = .282$, group (feigning vs. control), $F(2, 69) = 12.15, p < .001, \eta_p^2 = .261$, and an interaction between difficulty and group, $F(2, 70) = 10.31, p < .001, \eta_p^2 = .128$, for total hits. There was a decrease in the number of hits from Block 1 to 3 in the feigning group but not in the control group (see Table 1 and Figure 1). In contrast, there were no main or interaction effects for difficulty level or group for response time. For the group that was informed about the three levels of difficulty, there was a main effect for difficulty level (Block 1 to 3), $F(2, 36) = 3.80, p < .03, \eta_p^2 = .175$. Nonetheless, a nonsignificant pattern was observed in response time, $F(2, 37) = 2.74, p < .07, \eta_p^2 = .129$, in which only feigning participants took longer on the more "difficult" levels (see Table 1 and Figure 2).

In addition, analyses that were collapsed across all difficulty levels demonstrated that controls had significantly more hits than the feigning group, 29.35 versus 15.3 hits; $t(38) = 2.23, p = .002, d = 3.30$, as well as quicker response times, 1310 versus 2582 ms, respectively; $t(38) = -9.44, p = .000, d = .93$.

Cutoff, Sensitivity, and Specificity. In terms of the second objective, appropriate cutoff scores were established for hits (see Table 2). Cutoff scores were not created for response time due to the fact that this variable was not significant in distinguishing between the two conditions of the study (i.e., control vs. feigning). There was a high area under the curve for the total number of hits (.96), and the selected cutoff of ≥ 27 demonstrated a sensitivity of 95% and specificity of 95%.

Convergent Validity. CIH-EV total hits was highly correlated with hits on the second trial of the TOMM ($r = .897, n = 40, p = .001$), as well as the VSVT total responses ($r = .917, n = 40, p < .001$) and VSVT total time ($r = -.724, n$

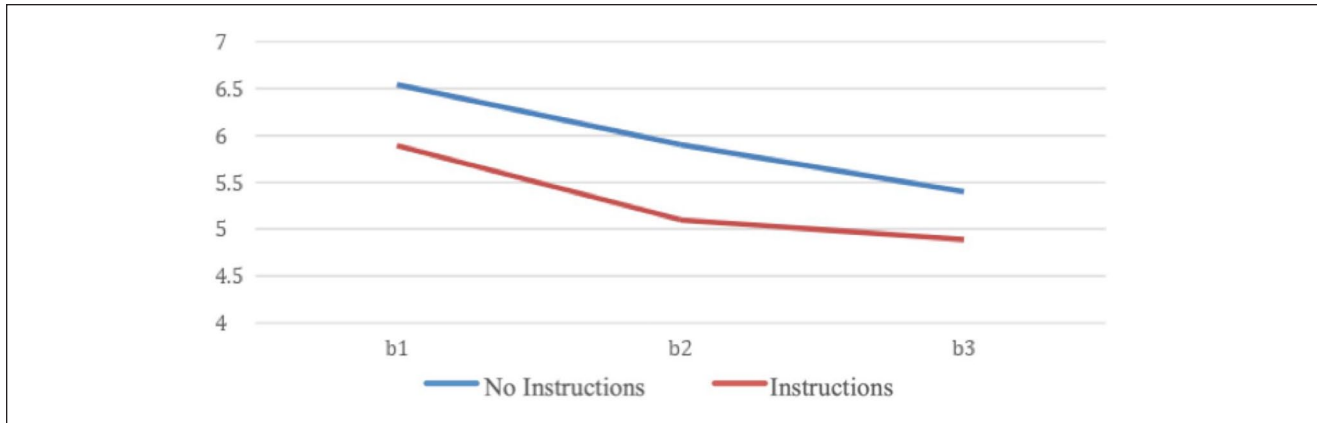


Figure 1. Number of correct responses for analog participants with and without instructions.
 Note. b = Block.

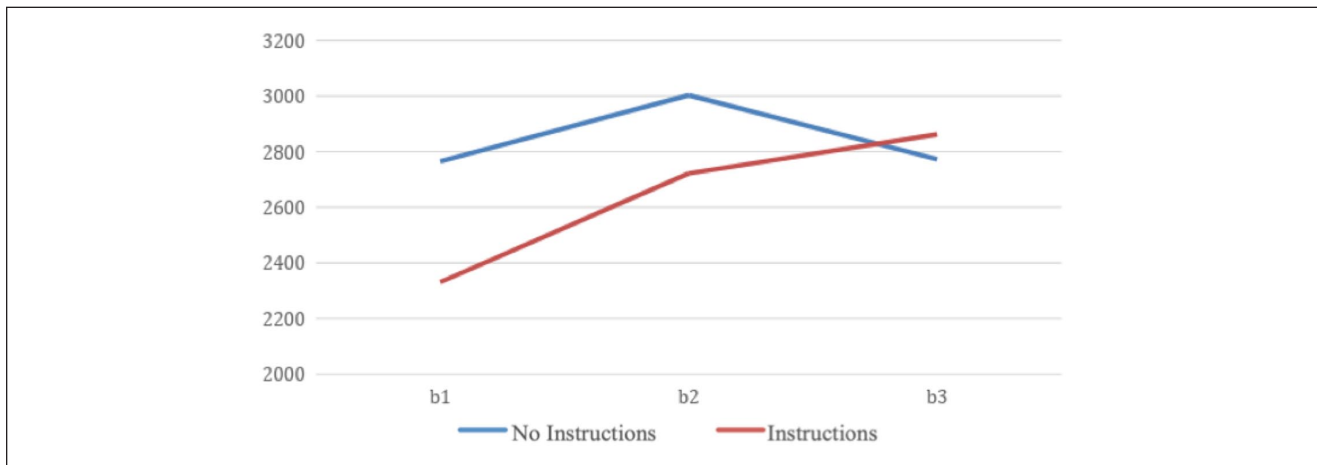


Figure 2. Response for analog participants with and without instructions.
 Note. b = Block.

Table 2. Sensitivity and Specificity CIH-EV Cutoff Scores for Spanish Sample.

Dependent variable	Blocks	Cutoff	Sensitivity	Specificity
Hits	1	≤7	0.789	0.95
		≤8	0.842	0.95
		≤9	0.895	0.9
	2	≤7	0.842	0.95
		≤8	0.847	0.95
		≤9	0.947	0.9
	3	≤7	0.895	0.95
		≤8	0.947	0.95
		≤9	0.947	0.9
Total	≤23	0.895	0.95	
	≤27	0.95	0.95	
	≤29	0.947	0.8	

= 40, $p = .005$). Last, we showed a moderate of congruence in the classification of feigning participants or controls between hits on the CIH-EV and the second trial of the TOMM ($\kappa = 1.00, p < .001$), and the VSVT ($\kappa = 1.00, p < .001$) and time ($\kappa = .547, p < .001$).

In sum, these results suggest that the CIH-EV was able to detect different levels of perceived difficulty for the feigning group by showing significantly fewer hits in this group. Response time, on the other hand, was not able to differentiate between the feigning and control groups. Furthermore, the CIH-EV performs with acceptable sensitivity and specificity, and showed a high level of agreement with other commonly used and validated forced-choice performance validity tests. Nonetheless, these specificity estimates should be cited with caution due to the fact that they were developed using a nonclinical sample of university students.

Study 2

Examining the Validity of the CIH-EV in Other Cultures

While the literature has shown that culture is associated with differences in neuropsychological performance (Puenta, Perez-Garcia, Lopez, Hidalgo-Ruzzante, & Fasfous, 2013), findings on the effect of culture in PVT are inconclusive. Only a small number of studies have been conducted on validity tests with culturally diverse and non-English speaking samples (Benuto & Leany, 2013; Weiss & Rosenfeld, 2012), with the majority of these studies focusing on Spanish-speaking samples (Gasquoine, Weimer, & Amador, 2017; Robles, Lopez, Salazar, Boone, & Glaser, 2015; Salazar, Lu, Wen, & Boone, 2007; Strutt, Scott, Shrestha, & York, 2011; Vilar-López et al., 2007). Furthermore, many of these studies combine Spanish speakers of different cultural backgrounds into the same group despite the fact that differences in validity testing have been found between different Spanish-speaking populations (Nijdam-Jones, Rivera, Rosenfeld, & Arango-Lasprilla, 2017). Due to these incomplete findings and the fact that, relative to the United States, there are few validated validity tests in other countries, more research is needed to understand how different cultural groups perform on forced-choice performance validity tests.

Therefore, the objective of second study was to assess the cultural equivalency of the CIH-EV using several languages and cultural contexts, including one English-speaking sample, one Portuguese-speaking sample, and two Spanish-speaking samples of distinct cultures. The instrument was translated and adapted into English and Portuguese, in addition to Spanish, as these are some of the most commonly spoken languages in the world (Paolillo & Das, 2006). It was also administered in two countries (Portugal and the United States), which do not share the same culture nor the same language (Portuguese and English). Furthermore, it was administered in a country (Colombia) that shares the same language (Spanish) but differs by culture with the country assessed in Study 1. We hypothesize that the CIH-EV will present similar cutoffs, sensitivity, and specificity in samples from the United States, Portugal, and Colombia.

Method

Participants. Participants from three different countries were recruited and randomly assigned to either the feigning or control group. At the Universidad del Norte (Barranquilla, Colombia), 42 total participants were recruited (24 female and 18 male), 21 of whom comprised the feigning group with an age range of 18 to 25 years ($M = 20.33$, $SD = 2.10$), and 21 who made up the control group with an age

range of 18 to 22 years ($M = 19.95$, $SD = 0.86$). In Portugal, 75 total participants (64 female and 11 male) were recruited from the Universidade de Lisboa, 38 of whom were assigned to the feigning group with an age range of 18 to 50 years ($M = 19.47$, $SD = 5.33$), and 37 in the control group with an age range between 17 and 33 years ($M = 19.02$, $SD = 3.05$). Finally, in the United States, 46 participants (26 female and 20 male) were recruited from Harvard University (Boston). There were 23 participants in the feigning group with ages ranging between 18 and 22 years ($M = 19.78$, $SD = 1.20$), and 23 in the control group between 18 and 22 years old ($M = 19.73$, $SD = 1.38$). As a part of inclusion criteria, it was required that participants' native and dominant language was the same as the language in which the test was administered. All participants in the United States took the performance validity tests in English, those in Portugal in Portuguese, and those in Colombia in Spanish. All participants were matched in education, having either just started or completed their undergraduate degree. Exclusion criteria were identical to those in Study 1.

Instruments. In all four countries (Spain, United States, Portugal, and Colombia), the CIH-EV and the TOMM were administered as these tests are available in all countries. A cutoff of 45 for the TOMM has been established for all four countries: Spain (Vilar-López et al., 2007), the United States (Tombaugh, 1996), Colombia (Puerta Lopera, Arango Tobón, Betancur Arias, & Sánchez Duque, 2016), and Portugal (Simões et al., 2017). For a full description of the CIH-EV and the TOMM, please see the instruments section of Study 1.

In the United States, the VSVT was administered in addition to the TOMM and CIH-EV. The suggested cutoff of 44 for the VSVT was applied to participants from the United States (Slick et al., 1997). This cutoff was chosen over the three-tiered cutoff established by Silk-Eglit, Lynch, and McCaffrey (2016) to compare findings with the Spanish population, for whom only a two-tiered cutoff has been established. In Colombia, the CIH-EV and the TOMM were the only tests administered. Finally, in Portugal, digit span of the Wechsler Adult Intelligence Scale Revised (WAIS-III; Wechsler, 2008) was administered in addition to the CIH-EV and TOMM. The WAIS-III Digit Span subtest consists of forward digit and backward digit span trials that become longer as the test continues. Test administration is discontinued when the participant fails a block of two sequences of the same length (Choi et al., 2014). The established cutoff scores of 5 for the total score of digit span, 4 for digits forward, and 2 for backward was applied to the Portuguese population (Castro, 2015; Pinho, 2012).

Procedure. All three studies were approved by the ethics committee in each location (i.e., Faculty of Psychology of the University of Lisbon, Universidad del Norte, and

Table 3. Selected CIH-EV Cutoff Scores for Spain (SP), Colombia (CO), Portugal (PT), and the United States (USA).

Dependent variable	Block	Cutoff				Sensitivity				Specificity			
		SP	CO	PT	USA	SP	CO	PT	USA	SP	CO	PT	USA
Hits	1	≤8	≤8	≤9	≤9.5	0.84	0.9	0.6	0.95	0.95	1	0.98	0.92
	2	≤8	≤8	≤8	≤8.5	0.84	0.95	0.62	0.91	0.95	1	1	0.96
	3	≤8	≤8	≤8	≤8.5	0.94	1	0.57	0.91	0.95	0.96	1	0.96
	Total	≤27	≤27	≤27	≤27	0.94	1	0.62	1	0.95	1	0.98	0.96

Note. CIH-EV = Coin in Hand–Extended Version.

Partners Healthcare Human Research Committee). Before test administration, participants signed an informed consent document and were informed about the voluntary nature of the study. Only participants in the United States received monetary compensation for participation, at \$10 each. Participants in Spain and Portugal received extra credit for their participation in their undergraduate courses. All participants were administered the CIH-EV in addition to other validity tests currently available and validated in each respective country: TOMM (Colombia); TOMM and the Digit Span subtest of the WAIS-III (Portugal); the TOMM and VSVT (the United States). All tests were administered in a randomized order, and test administrators were blind to the participants' condition (i.e., feigning or control).

As we have found a trend toward a greater differentiation in performance in time for feigning participants when using instructions (see Table 1 and Figure 2), specific instructions about the various levels of difficulty were given to all participants, both feigning and control. All participants performed validity tests in a quiet and isolated room. The same protocol that was outlined in Study 1 regarding the feigning comparison group developed by Suhr and Gunstad (2000) was replicated in these three samples.

Design and Data Analyses. A cutoff score was established for correct hits and time on the CIH-EV for each country applying the same methodology as described in Study 1. Cutoffs for response time, on the other hand, were not established due to the fact that they were not significant in detecting feigning in Study 1. Using these new cutoffs, chi-square analyses were conducted for each country to measure convergent validity by assessing the rates of false positives and negatives. A cutoff score of 45 was applied to Trial 2 of the TOMM, which has been validated in Spain (Vilar-López et al., 2007), Colombia (Puerta Lopera et al., 2016), the United States (Tombaugh, 1996), and Portugal (Mota et al., 2008). As for the VSVT, a cutoff score of 44 of was applied to both Spain (Vilar-López et al., 2007) and the United States (Slick et al., 1997). Finally, a cutoff of 5 for the total score of digit span, 4 for digits forward, and 2 for backward was applied to scores for the Portuguese population (Castro, 2015; Pinho, 2012). In addition, Spearman correlations were employed to test the association between scores on various validity tests.

Table 4. Sensitivity and Specificity CIH-EV Cutoff Scores for U.S. Population.

Dependent variable	Blocks	Cutoff	Sensitivity	Specificity
Hits	1	≤8	0.739	1
		≤9	0.957	0.913
		≤10	1	0
	2	≤7	0.652	1
		≤8	0.91	0.957
		≤9	0.957	0.87
		≤10	1	0
	3	≤7	0.652	1
		≤8	0.913	0.957
		≤9	0.957	0.87
		≤10	1	0
	Total	≤27	0.957	1
		≤28	1	0.957
≤29		1	0.783	
≤30		1	0	

Note. CIH-EV = Coin in Hand–Extended Version.

Results

The United States: Cutoff, Sensitivity, Specificity, and Convergent Validity. Cutoffs for hits (≥ 27) were selected (Table 3), and results showed a high area under the curve (.999) for the U.S. sample. For a full list of cutoff scores and respective specificity and sensitivity, please refer Table 4.

Portugal: Cutoff, Sensitivity, Specificity, and Convergent Validity. Cutoffs for hits (27) were selected (see Table 3), and results showed a high area under the curve (.807) for the Portuguese sample. Cutoff scores and respective specificity and sensitivity can be referred in Table 5.

Colombia: Cutoff, Sensitivity, Specificity, and Convergent Validity. Cutoffs for hits (27) were selected (see Table 3), and results showed a high area under the curve (1) for the Colombian sample. Cutoff scores and respective specificity and sensitivity can be referenced in Table 6.

Spearman correlations revealed a positive correlation between the number of hits on the CIH-EV and TOMM Trial 2 in the Colombian ($r = .868, n = 42, p < .001$), Portuguese ($r = .671, n = 75, p < .001$), and U.S. sample

Table 5. Sensitivity and Specificity CIH-EV Cutoff Scores for the Portuguese Sample.

Dependent variable	Blocks	Cutoff	Sensitivity	Specificity
Hits	1	≤8	0.543	1
		≤9	0.6	0.973
		≤10	1	0
	2	≤7	0.543	1
		≤8	0.629	1
		≤9	0.629	0.946
		≤10	1	0
	3	≤8	0.543	1
		≤8	0.571	1
		≤9	0.686	0.865
	Total	≤25	0.6	1
		≤27	0.629	1
		≤28	0.629	0.973
≤29		0.686	0.811	

Note. CIH-EV = Coin in Hand–Extended Version.

Table 6. Sensitivity and Specificity CIH-EV Cutoff Scores for the Colombian Population.

Dependent variable	Blocks	Cutoff	Sensitivity	Specificity
Hits	1	≤6	0.619	1
		≤8	0.905	1
		≤9	0.905	0.905
		≤10	1	0
	2	≤7	0.857	1
		≤8	0.952	1
		≤9	1	0.81
	3	≤7	0.905	1
		≤8	1	0.952
		≤9	1	0.857
	Total	≤26	0.952	1
		≤27	1	1
		≤28	1	0.857

Note. CIH-EV = Coin in Hand–Extended Version.

($r = .853$, $n = 46$, $p < .001$). In the Portuguese sample, there was a positive correlation between the total number of hits on the CIH-EV and the Digit Span subtest of the WAIS-III ($r = .493$, $n = 75$, $p < .001$). In the U.S. sample, there was a positive correlation between the VSVT and the CIH-EV for both time ($r = -.704$, $n = 46$, $p < .001$) and hits ($r = .830$, $n = 46$, $p < .001$).

Last, chi-square analysis demonstrated that there was a high degree of congruence between hits on the CIH-EV and the second trial of the TOMM in Colombia ($\kappa = .952$, $p < .001$], Portugal ($\kappa = .750$, $p < .001$), and the United States ($\kappa = .737$, $p < .001$). Furthermore, there was an association between the total number of hits on the CIH-EV and Digit Span subtest ($\kappa = .345$, $p < .001$) in

the Portuguese sample. In the U.S. sample, there was also a high degree of congruence between the CIH-EV and the VSVT for both hits ($\kappa = .913$, $p < .001$) and time ($\kappa = .953$, $p < .001$). All of these correlations support the validity of the CIH-EV, showing its ability to perform at a level comparable to other commonly used performance validity tests.

Discussion

The objective of these studies was to validate and test the cultural equivalency of a modified version of the CIH-EV, a computerized and visual performance validity test. The CIH-EV demonstrated a specificity of at least 95% and a sensitivity of at least 62%, resulting in an observed rate of 5% false positives. Performance on the CIH-EV was strongly correlated with other validity tests, such as the TOMM, VSVT, and Digit Span subtest of the WAIS. Finally, there was high cultural equivalence across countries, which was reflected in the same cutoff score that was established for countries that varied by both culture and language. To our knowledge, this is the first royalty-free computerized performance validity test to demonstrate these positive psychometric properties across different languages and cultures.

Study 1 demonstrated positive results for the modifications of the CIH-EV. With the inclusion of different levels of perceived difficulty, results showed that feigning (i.e., faking bad) participants inferred varying levels of difficulty which was reflected in poorer performance on the “difficult” levels. Performance of the control participants, on the other hand, did not differ by difficulty level. These results suggest that this modification made to the original CIH may help further distinguish feigning from nonfeigning individuals. However, our findings revealed that there were no significant differences between the conditions of being given instructions or not about the levels of difficulty. Nonetheless, a pattern was found among feigning participants such that those informed about the incrementing levels of difficulty had longer response times, in comparison with those who were not informed. While these differences did not reach significance, this response pattern suggests that having knowledge about the difficulty levels may influence performance to some degree. In line with previous research, some studies have shown that including multiple levels of perceived difficulty can improve the detection of simulation (Bickart, Meyer, & Connell, 1991; Chiu & Lee, 2002; Slick, Hopp, Strauss, Hunter, & Pinch, 1994). Due to these findings and support from the literature, we recommend offering instructions about the seemingly increasing difficulty prior to CIH-EV administration in order to promote greater differentiation between simulating and nonsimulating individuals.

In terms of the properties of the CIH-EV in Study 1, the test performed with high sensitivity (94%) and specificity

(95%). The obtained rates surpass the criteria of at least 90% specificity and 50% sensitivity as outlined by Sugarman and Axelrod (2015). These findings suggest that the CIH-EV is successfully able to distinguish between feigning and control participants without risking high rates of false positives among nonclinical individuals. Other studies conducted on the original CIH (Kapur, 1994) have similarly found low rates of false positives. In fact, those conducted on feigning participants found 0% false positive error rates when using a cutoff of >1 (Cochrane et al., 1998; Hanley et al., 1999; Kelly et al., 2005; Schroeder, Peck, et al., 2012). Our findings also demonstrate that the digitization and translation of the CIH measure did not diminish the capacity of the test to detect feigning. Similar to the high sensitivity and specificity found in Study 1, high sensitivity and specificity were also found in the United States, Portugal, and Colombia. However, it should be emphasized that these findings were gathered on healthy nonclinical individuals who were asked to exaggerate cognitive impairment. The sensitivity and specificity estimates obtained in this study are likely to be inflated and are likely higher than those that would be found in clinical samples or individuals who suffer genuine deficits. As such, estimates should not be applied to other populations due to the increased risk of committing false positive errors and the immense forensic and clinical implications doing so may have. To this end, the estimates presented here should be considered a starting point in the validation of the CIH-EV, and clinicians should wait to apply this instrument to clinical populations until it is cross-validated among genuine clinical patients. These tests were designed to serve as analog studies, so future research with actual cognitive impairment is necessary.

In addition to finding low false positive rates, we found that the CIH-EV demonstrates excellent convergent validity. Performance on the CIH-EV was highly correlated with that of other commonly used performance validity measures, such as the VSVT, TOMM, and Digit Span subtest of the WAIS. Furthermore, there was a significant overlap between the CIH-EV and these three performance validity tests in their ability to categorize feigning and nonfeigning individuals. Comparable performance between these measures supports the use of the CIH-EV as a valid alternative to other commonly used performance validity tests.

In terms of cross-cultural comparisons, identical cutoff scores for hits were obtained for all countries. Although there were different cutoffs for time, this variable was not considered in cross-cultural comparisons as it was not as good of a predictor as was hits in detecting feigning. Nonetheless, in each country, the response times do correlate with whether the participant is feigning, with simulators exhibiting generally longer response times. A similar finding was obtained previously with the VSVT, where time was found to have less utility in comparison with error scores in detecting feigning but was insignificantly correlated with the participant's condition (Slick et al., 1997).

To the best of our knowledge, this is also the first study to compare a computerized performance validity test across three different languages. The present study did not find differences between the four countries, which included two samples that speak the same language but have different cultures (Colombia and Spain). These findings suggest the absence of cross-cultural effects on CIH-EV test performance. Additionally, in line with other research conducted on culturally different samples using forced-choice performance validity tests, Vilar-López et al. (2007) demonstrated that performance on the TOMM and VSVT was almost identical between North American and Spanish samples, and that the same cutoff scores could be applied. The fact that the CIH employs visual stimuli as opposed to verbal may help make the instrument less language dependent than other performance validity tests such as Digit Span of the WAIS. This feature is imperative for individuals who are not literate or who have diminished verbal capacities. While more research is needed to test the cross-cultural and linguistic applications of the CIH-EV, the present study has taken a step in this direction by establishing specific cutoff scores for each language and culture.

Limitations

A major limitation of the present study is its generalizability to other populations, such as individuals in clinical and forensic settings. While knowing the sensitivity and specificity of college students provides some useful information, it is imperative to use specificity and sensitivity estimates that have been specifically developed for each type of clinical group. Due to the fact that the present study employed young and neurologically healthy nonclinical college students, the false positive rates are very low. For these reasons, caution must be taken when interpreting these results. Future research may consider applying the CIH-EV to clinical populations, such as individuals with true cognitive impairment due to traumatic brain injury or dementia. This analog design was chosen for the preliminary validation of the CIH-EV as it allows for the experimental comparison of feigning and nonfeigning groups. It is imperative for future studies to test this instrument on individuals with true cognitive impairment in a nonhighly educated sample as well as populations with a high probability of feigning in order to determine whether genuine cognitive impairment or litigation involvement affect its predictive capacity. In addition to these shortcomings, there may be certain complications associated with the computerized application of the CIH-EV. Facilities with financial restrictions may have difficulties purchasing a tablet or computer, and others may have limited access to such modalities due to heightened security, such as in correctional settings.

In sum, this work represents the preliminary validation of the first computerized performance validity test in the public domain. Offering the CIH-EV in a digitized manner may

allow for more uniformity across studies in the future in terms of instrumentation and methodology. Its computerized format also allows for an improved standardization and precision of data collection and recording of variables, such as response time, that are more difficult to measure in traditional pen-and-paper performance validity tests. Furthermore, it may help encourage the use of a wider range of instruments in validity testing, which has been strongly recommended by researchers and clinicians alike to address issues associated with coaching and familiarization with tests (Heilbronner et al., 2009; Martin et al., 2015; Ruocco et al., 2008; Schutte, Millis, Axelrod, & VanDyke, 2011; Van Dyke, Millis, Axelrod, & Hanks, 2013). Finally, the availability of the test in three different languages will help extend access to larger populations internationally. The CIH-EV is currently available on request for credentialed psychologists (please contact the first author).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a predoctoral fellowship of the Ministry of Education and Professional Training in Spain. Funding was also received by authors at Harvard University from the Harvard College Research Program to cover costs for participant collaboration.

ORCID iDs

Julia C. Daugherty  <https://orcid.org/0000-0002-1221-4394>

Carlos Jose De los Reyes-Aragon  <https://orcid.org/0000-0003-4472-0635>

References

- Benuto, L. T., & Leany, B. D. (2013). Assessing effort and malingering with the Hispanic client. In L. T. Benuto (Ed.), *Guide to psychological assessment with Hispanics* (pp. 99-112). New York, NY: Springer Science Business Media. doi:10.1007/978-1-4614-4412-1_7
- Bickart, W. T., Meyer, R. G., & Connell, D. (1991). The symptom validity technique as a measure of feigned short-term memory deficit. *American Journal of Forensic Psychology, 9*, 3-11.
- Binder, L. M. (1990). Malingering following minor head trauma. *The Clinical Neuropsychologist, 4*, 25-36.
- Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-item recognition trial: A technique to enhance sensitivity of the Rey 15-item memorization test. *Journal of Clinical and Experimental Neuropsychology, 24*, 561-573.
- Boone, K. B., Victor, T. L., Wen, J., Razani, J., & Ponton, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology, 22*, 355-365. doi:10.1016/j.acn.2007.01.010
- Castro, V. S. S. (2015). *Deteção da simulação em indivíduos com e sem diagnóstico de depressão* [Simulation Detection in Individuals with and without depression diagnosis]. Master's thesis. Universidade de Aveiro, Aveiro, Portugal. Retrieved from <https://ria.ua.pt/handle/10773/15613>
- Chafetz, M. (2011). Reducing the probability of false positives in malingering detection of social security disability claimants. *The Clinical Neuropsychologist, 25*, 1239-1252. doi:10.1080/13854046.2011.586785
- Chiu, V. W., & Lee, T. (2002). Detection of malingering behavior at different levels of task difficulty in Hong Kong Chinese. *Rehabilitation Psychology, 47*, 194-203. doi:10.1037/0090-5550.47.2.194
- Choi, H. J., Lee, D. Y., Seo, E. H., Jo, M. K., Sohn, B. K., Choe, Y. M., . . . Jhoo, J. H. (2014). A normative study of the digit span in an educationally diverse elderly population. *Psychiatry Investigation, 11*, 39-43. doi:10.4306/pi.2014.11.1.39
- Cochrane, H. J., Baker, G. A., & Meudell, P. R. (1998). Simulating a memory impairment: Can amnesics implicitly outperform simulators? *British Journal of Clinical Psychology, 37*, 31-48. doi:10.1111/j.2044-8260.1998.tb01277.x
- Colwell, K., & Sjerven, E. R. (2005). The "coin-in hand" stratagem for the forensic assessment of malingering. *American Journal of Forensic Psychology, 23*, 83-90.
- Daugherty, J. C., Hidalgo-Ruzzante, N., & Pérez-García, M. (2017). Coin @ Hand: Project Believe [Mobile Application Software]. Retrieved from <http://itunes.apple.com>
- Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist, 22*, 705-722. doi:10.1080/1385404701440493
- Ferreira, I., Gomes, F., Moreira, I., Silva, A., & Cavaco, S. (2017). Medir a validade da avaliação neuropsicológica em doentes com esclerose múltipla ou lúpus eritematoso sistémico [Measuring the validity of neuropsychological evaluations in patients with multiple sclerosis or systemic lupus erythematosus]. *Revista Evidências, 2017*, 27.
- Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment, 14*, 3-11. doi:10.1177/1073191106292009
- Gasquoine, P. G., Weimer, A. A., & Amador, A. (2017). Specificity rates for non-clinical, bilingual, Mexican Americans on three popular performance validity measures. *The Clinical Neuropsychologist, 31*, 587-597. doi:10.1080/13854046.2016.1277786
- Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., & Wagster, M. V. (2010). Assessment of neurological and behavioural function: The NIH Toolbox. *Lancet Neurology, 9*, 138-139. doi:10.1016/s1474-4422(09)70335-7
- Griffin, G. A. E., Glassmire, D. M., Henderson, E. A., & McCann, C. (1997). Rey II: Redesigning the Rey screening test of malingering. *Journal of Clinical Psychology, 53*, 757-766. doi:10.1002/(SICI)1097-4679(199711)53:7<757::AID-JCLP14>3.0.CO;2-K
- Hampson, N. E., Kemp, S., Coughlan, A. K., Moulin, C. J., & Bhakta, B. B. (2014). Effort test performance in clinical acute brain injury, community brain injury, and epilepsy popula-

- tions. *Applied Neuropsychology: Adult*, 21, 183-194. doi:10.1080/09084282.2013.787425
- Hanley, J. R., Baker, G. A., & Ledson, S. (1999). Detecting the faking of amnesia: A comparison of the effectiveness of three different techniques for distinguishing simulators from patients with amnesia. *Journal of Clinical and Experimental Neuropsychology*, 21, 59-69. doi:10.1076/jcen.21.1.59.936
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093-1129. doi:10.1080/13854040903155063
- Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, 11, 967-974. doi:10.1080/01688638908400949
- Iverson, G. L., Franzen, M. D., & McCracken, L. M. (1994). Application of a forced-choice memory procedure designed to detect experimental malingering. *Archives of Clinical Neuropsychology*, 9, 437-450. doi:10.1093/arclin/9.5.437
- Kapur, N. (1994). The coin-in-the-hand test: A new "bed-side" test for the detection of malingering in patients with suspected memory disorder. *Journal of Neurology, Neurosurgery, & Psychiatry*, 57, 385. doi:10.1136/jnnp.57.3.385
- Kelly, P. J., Baker, G. A., Broek, M. D., Jackson, H., & Humphries, G. (2005). The detection of malingering in memory performance: The sensitivity and specificity of four measures in a UK population. *British Journal of Clinical Psychology*, 44, 333-341. doi:10.1348/014466505X35687
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410-425. doi:10.1076/clin.17.3.410.18089
- Larrabee, G. J. (2012). Assessment of malingering. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed., pp. 116-159). New York, NY: Oxford University Press.
- Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334-371). New York, NY: Oxford University Press.
- Lezak, M. D. (2004). *Neuropsychological assessment*. New York, NY: Oxford University Press.
- Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, 32, 391-421.
- Martin, P. K., Schroeder, R. W., Heinrichs, R. J., & Baade, L. E. (2015). Does true neurocognitive dysfunction contribute to Minnesota Multiphasic Personality Inventory-2nd Edition-Restructured Form cognitive validity scale scores? *Archives of Clinical Neuropsychology*, 30, 377-386. doi:10.1093/arclin/acv032
- Merten, T., Green, P., Henry, M., Blaskewitz, N., & Brockhaus, R. (2005). Analog validation of German-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology*, 20, 719-726. doi:10.1016/j.acn.2005.04.004
- Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, 18, 261-276. doi:10.1093/arclin/18.3.261
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094-1102.
- Morgan, J. E., & Sweet, J. J. (2009). *Neuropsychology of malingering casebook*. New York, NY: Psychology Press.
- Mota, M., Simões, M. R., Amaral, L., Dias, I., Luís, D., Pedrosa, C., . . . Silva, I. (2008). Test of Memory Malingering (TOMM): Estudos de validação numa amostra de reclusos [TOMM: Validation studies on a sample of inmates]. *Psiquiatria, Psicologia & Justiça*, 2, 23-41.
- Nijdam-Jones, A., Rivera, D., Rosenfeld, B., & Arango-Lasprilla, J. C. (2017). A cross-cultural analysis of the Test of Memory Malingering among Latin American Spanish-speaking adults. *Law and Human Behavior*, 41, 422-428. doi:10.1037/lhb0000250
- Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*, 29, 1321-1336. doi:10.1037/pas0000438
- Paolillo, J. C., & Das, A. (2006). *Evaluating language statistics: The ethnologue and beyond*. Retrieved from http://uis.unesco.org/sites/default/files/documents/evaluating-language-statistics-the-ethnologue-and-beyond-en_0.pdf
- Pinho, J. I. D. S. S. (2012). *Testes de validade de sintomas: Validação de um protocolo em amostras de adultos idosos* (Symptom Validity Tests: Validation of a protocol in a sample of senior adults, Master's thesis, Universidade de Coimbra, Coimbra, Portugal). Retrieved from <https://estudogeral.sib.uc.pt/handle/10316/23453>
- Puente, A. E., Perez-Garcia, M., Lopez, R. V., Hidalgo-Ruzzante, N. A., & Fasfous, A. F. (2013). Neuropsychological assessment of culturally and educationally dissimilar individuals. In *Handbook of multicultural mental health* (2nd ed., pp. 225-241). San Diego, CA: Academic Press.
- Puerta Lopera, I. C., Arango Tobón, E., Betancur Arias, J. D., & Sánchez Duque, J. W. (2016). Diagnóstico de Simulación de Disfunción Neurocognitiva en sujetos que presentan accidentes laborales [Simulation diagnostics of neurocognitive dysfunction in subjects that present occupational accidents]. *Archivos de Medicina*, 16, 445-454.
- Rey, A. (1964). *L'examen clinique en psychologie* [The Clinical Psychological Evaluation]. Paris, France: Presses Universitaires de France.
- Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury*, 19, 539-543. doi:10.1080/02699050400005242
- Robles, L., Lopez, E., Salazar, X., Boone, K. B., & Glaser, D. F. (2015). Specificity data for the b Test, Dot Counting Test, Rey-15 Item Plus Recognition, and Rey Word Recognition Test in monolingual Spanish-speakers. *Journal of Clinical and Experimental Neuropsychology*, 37, 614-621. doi:10.1080/13803395.2015.1039961
- Rudman, N., Oyeboode, J. R., Jones, C. A., & Bentham, P. (2011). An investigation into the validity of effort tests in a working

- age dementia population. *Aging & Mental Health*, 15, 47-57. doi:10.1080/13607863.2010.508770
- Ruocco, A. C., Swirsky-Sacchetti, T., Chute, D. L., Mandel, S., Platek, S. M., & Zillmer, E. A. (2008). Distinguishing between neuropsychological malingering and exaggerated psychiatric symptoms in a neuropsychological setting. *The Clinical Neuropsychologist*, 22, 547-564. doi:10.1080/13854040701336444
- Salazar, X. F., Lu, P. H., Wen, J., & Boone, K. B. (2007). The use of effort tests in ethnic minorities and in non-English speaking and English as a second language populations. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 405-427). New York, NY: Guilford Press.
- Schroeder, R. W., Peck, C. P., Buddin, W. H., Jr., Heinrichs, R. J., & Baade, L. E. (2012). The Coin-in-the-Hand Test and dementia: More evidence for a screening test for neurocognitive symptom exaggeration. *Cognitive and Behavioral Neurology*, 25, 139-143. doi:10.1097/wnn.0b013e31826b71c1
- Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable Digit Span: A systematic review and cross-validation study. *Assessment*, 19, 21-30. doi:10.1177/1073191111428764
- Schutte, C., Millis, S. R., Axelrod, B. N., & VanDyke, S. (2011). Derivation of a composite measure of embedded symptom validity indices. *The Clinical Neuropsychologist*, 25, 454-462. doi:10.1080/13854046.2010.550635
- Silk-Eglit, G. M., Lynch, J. K., & McCaffrey, R. J. (2016). Validation of Victoria Symptom Validity Test cutoff scores among mild traumatic brain injury litigants using a known-groups design. *Archives of Clinical Neuropsychology*, 31, 231-245. doi:10.1093/arclin/acv108
- Simões, M. R., Sousa, L. B., Fonseca, M. S., Garcia, S. G., Pinho, J. I., Soares, D. L., . . . Pinho, M. S. (2017). Teste de Simulação de Problemas de Memória (Test of Memory Malingering, TOMM). In M. R. Simões, L. S. Almeida, & M. Gonçalves (Eds.), *Psicologia Forense: Instrumentos de Avaliação* [Forensic Psychology: Evaluation Instruments] (pp. 353-375). Lisbon, Portugal: Pactor.
- Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's Forced-Choice Memory Test. *Journal of Clinical and Experimental Neuropsychology*, 16, 472-481. doi:10.1080/01688639408402657
- Slick, D., Hopp, G., Strauss, E., & Thompson, G. (1997). *The Victoria Symptom Validity Test*. Odessa, FL: PAR.
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545-561. doi:10.1076/1385-4046(199911)13:04;1-Y;FT545
- Sollman, M. J., Ranssen, J. D., & Berry, D. T. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, 22, 325-335. doi:10.1037/a0018857
- Stimmel, M., Green, D., Belfi, B., & Klaver, J. (2012). Exploring the accuracy and utility of the Rey Fifteen Item Test (RMT) with recognition trial in a forensic psychiatric population. *International Journal of Forensic Mental Health*, 11(1), 51-58. doi.org/10.1080/14999013.2012.676148
- Strutt, A. M., Scott, B. M., Shrestha, S., & York, M. K. (2011). The Rey 15-Item Memory Test and Spanish-speaking older adults. *The Clinical Neuropsychologist*, 25, 1253-1265. doi:10.1080/13854046.2011.609839
- Sugarman, M. A., & Axelrod, B. N. (2015). Embedded measures of performance validity using verbal fluency tests in a clinical sample. *Applied Neuropsychology: Adult*, 22, 141-146. doi:10.1080/23279095.2013.873439
- Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, 15, 415-424. doi:10.1093/arclin/15.5.415
- Suhr, J. A., & Gunstad, J. (2007). Coaching and malingering: A review. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 287-311). New York, NY: Oxford University Press.
- Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. New York, NY: MHS.
- Tydecks, S., Merten, T., & Gubbay, J. (2006). The Word Memory Test and the One-in-Five-Test in an analogue study with Russian-speaking participants. *International Journal of Forensic Psychology*, 1, 29-37.
- Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist*, 27, 1234-1246. doi:10.1080/13854046.2013.835447
- Vilar-López, R., Santiago-Ramajo, S., Gómez-Río, M., Verdejo-García, A., Llamas, J. M., & Pérez-García, M. (2007). Detection of malingering in a Spanish population using three specific malingering tests. *Archives of Clinical Neuropsychology*, 22, 379-388. doi:10.1016/j.acn.2007.01.012
- Wechsler, D. (2008). *Escala de Inteligência de Wechsler para Adultos: 3ª edição [Wechsler Adult Intelligence Scale: 3rd edition]. Manual 1ª Edição Portuguesa [1st Portuguese Edition]*. Lisbon, Portugal: CEGOC-TEA.
- Weiss, R. A., & Rosenfeld, B. (2012). Navigating cross-cultural issues in forensic assessment: Recommendations for practice. *Professional Psychology: Research and Practice*, 43, 234-240. doi:10.1037/a0025850
- Whitney, K. A., Hook, J. N., Steiner, A. R., Shepard, P. H., & Callaway, S. (2008). Is the Rey 15-Item Memory Test II (Rey II) a valid symptom validity test? Comparison with the TOMM. *Applied Neuropsychology*, 15, 287-292. doi:10.1080/09084280802325215
- Yeh, T. C., Chou, Y. C., Weng, J. P., Yeh, H. W., Kao, Y. C., Chiang, W. S., & Tzeng, N. S. (2018). Detection of malingering in the memory of patients with dementia: A pilot study on coin-in-the-hand test in a Northern Taiwan Memory Clinic. *Journal of Medical Sciences*, 39, 81-89. doi:10.4103/jmedsci.jmedsci_100_18
- Youngjohn, J. R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment*, 2, 279-283. doi:10.1177/1073191195002003007